# A METHOD FOR EVALUATING IMPROPER PRIOR DISTRIBUTIONS

by

Morris L. Eaton[1]
University of Minnesota

Technical Report No. 389
June 1981

AMS 1970 Subject Classification: Primary-62C05; Secondary 62C10, 62C15.

Key Words and Phrases: Posterior distributions, Improper priors, Bayes rules, Fair scoring functions, admissibility.

[1]This work was supported in part by National Science Foundation Grant MCS77-25112.

## Abstract

In this paper, posterior distributions are interpreted as decision functions. Using this interpretation, it is proposed that improper prior distributions be evaluated in a decision theoretic manner by evaluating the posterior distributions they define. It is argued that the loss structure for the problem should produce Bayes rules which are simply the posterior distributions. Examples of such loss functions are given and admissibility is discussed.

## §1: Introduction

In this paper, we explore a formulation of a decision problem which, among other things, allows the evaluation of improper prior distributions via the posterior distributions they define. This formulation arose from an attempt to decide whether or not some of the classical fiducial distributions were in a decision theoretic sense "reasonable" (e.g., admissible, minimax) decision rules. For example, if $X$ given $\theta$ is $N(\theta,1)$, then the pivotal $X-\theta$ leads to the fiducial distribution "$\theta$ given $X$ is $N(X,1)$". Within the Bayesian framework, this posterior distribution for $\theta$ arises by assuming $\theta$ has the improper prior distribution $d\theta$ (Lebesgue measure) on $R^1$. In the traditional estimation problem (e.g., estimating $\theta$ with quadratic loss) it is ordinarily the case that the non-randomized estimators form an essentially complete class. Thus, the randomized decision rule "$\theta$ given $X$ is $N(X,1)$" cannot be reasonable in a traditional formulation of the estimation problem where decision rules are compared via their risk functions. This situation led to a re-examination of the decision theoretic formulation of the estimation problem and resulted in the reformulation of such problems which we will now briefly describe.

Consider a sample space $(\mathcal{X},\mathcal{F})$ and a parameter space $(\Theta,\mathcal{B})$ where $\mathcal{B}$ is some natural $\sigma$-algebra of subsets of $\Theta$. Let $\mathcal{M}(\mathcal{B})$ be the set of all probability measures defined on $\mathcal{B}$. Consider an observable $X \in \mathcal{X}$ whose distribution, given $\theta$, is one of a family $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$. By an <u>inference</u> we mean a function $\delta$ mapping $\mathcal{X}$ into $\mathcal{M}(\mathcal{B})$ - that is, for each $x$, $\delta(\cdot|x)$ is a probability distribution defined on $(\Theta,\mathcal{B})$. A similar point of view was also taken in Bernardo (1979). This is just the usual definition of a randomized decision rule (at this point all measurability issues will be ignored). In the language of decision theory, $M(\mathcal{B})$ is the action space. Suppose a loss function $L$ defined on $M(\mathcal{B}) \times \Theta$ to $[0,\infty)$ is specified; $L(\nu,\theta)$ measures the loss for action $\nu \in M(\mathcal{B})$ when $\theta$ is the "state of nature". The <u>risk function of $\delta$</u> is defined by

$$R(\delta,\theta) = \int_{X} L(\delta(\cdot|x),\theta)P(dx|\theta)$$

for any inference $\delta$. The domain of definition of the risk function is extended to $M(\mathcal{B})$ by

$$\bar{R}(\delta,\pi) = \int R(\delta,\theta)\pi(d\theta), \quad \pi \in M(\mathcal{B}) .$$

Now, we ask "What properties should $R$ have?". First suppose a Bayesian has a prior distribution $\pi$ which represents beliefs about $\theta$. The rules of probability dictate the Bayesian's inference about $\theta$ after seeing $X$ - namely, the Bayesian's inference is just the posterior distribution calculated from the model $\mathcal{P} = \{P(\cdot|\theta)|\theta \in \Theta\}$ and the prior $\pi$. Let $Q_{\pi}(\cdot|x)$ be the calculated posterior distribution when $X = x$ so that $Q_{\pi}$ is an inference. In order that our formulation lead the Bayesian to the inference $Q_{\pi}$ (via minimizing the average risk) the risk function must satisfy

$$(1.1) \qquad \bar{R}(\delta,\pi) \geq \bar{R}(Q_{\pi},\pi)$$

for all inferences $\delta$ and $\pi \in M(\mathcal{B})$. To pinpoint the issues involved, let us summarize the arguments which led to (1.1).

(i) Given $X = x$, an inference about $\theta$ is a probability distribution on $\Theta$ (supposedly reflecting our knowledge about $\theta$ based on $X$ and whatever else we know)

(ii) If a Bayesian's prior opinion is given by $\pi$, then the Bayesian's inference must by $Q_\pi$

(iii) In order that our formulation of the decision problem be consistent with the accepted methods of updating prior information, inequality (1.1) must hold.

The obvious question is whether or not such R's exist. First observe that the space $\mathcal{X}$ played a secondary role in the above argument. In particular, if $\mathcal{X}$ consists of only one point (no information about $\theta$ is contained in X) then (1.1) becomes

$$(1.2) \qquad \int L(\nu,\theta)\pi(d\theta) \geq \int L(\pi,\theta)\pi(d\theta), \quad \pi,\nu \in \mathcal{M}(\mathcal{B}) .$$

(Here, we have replaced $\delta$ by $\nu$ in (1.1) and suppressed the dependence on x since there is only one x). Hence, if (1.1) is to hold for all experiments, then (1.2) must hold. Conversely, if (1.2) holds, then

$$(1.3) \qquad \bar{R}(\delta,\pi) = \int_\Theta \int_{\mathcal{X}} L(\delta(\cdot|x),\theta)P(dx|\theta)\pi(d\theta) =$$

$$\int_{\mathcal{X}} \int_\Theta L(\delta(\cdot|x),\theta)Q_\pi(d\theta|x)m_\pi(dx)$$

where $m_\pi$ is the marginal distribution of X when the prior is $\pi$. Using (1.2) (with $\nu = \delta(\cdot|x)$ and $\pi$ replaced by $Q_\pi(\cdot|x)$) on the right most member of (1.3) yields

(1.4) $$\bar{R}(\delta,\pi) \geq \int_X\int_\Theta L(Q_\pi(\cdot|x),\theta)Q_\pi(d\theta|x)m_\pi(dx) =$$

$$\int_\Theta\int_X L(Q_\pi(\cdot|x),\theta)P_\theta(dx)\pi(d\theta) = \bar{R}(Q_\pi,\pi)$$

which is just (1.1).

Loss functions which satisfy criteria (1.2) have arisen in a number of other contexts. For example, Brier (1950) introduced such a criteria in scoring weathermen on their weather forecasts and Good (1952) indepe-pendently introduced this criteria in a more general context. For a survey and an extensive bibliography dealing primarily with the case when $\Theta$ is a finite set, see Savage (1971). For $\Theta$ infinite when densities are assumed, see Hendrickson and Buehler (1971). In Section 2, we will describe a class of L's which satisfy (1.2) in the general context described above.

Since it was a Bayesian argument which led to (1.2), L's which satisfy (1.2) will be called fair Bayes loss functions (FBLF). Further, a decision problem with the property that a Bayesian's decision rule is just the Bayesian's posterior distribution will be called a fair Bayes decision problem (FBDP). For the remainder of this paper, all decision problems will be FBDP's.

Given a FBDP and a prior $\pi$, then the appropriate inference (decision rule) is the posterior $Q_\pi$. Under fairly mild regularity conditions, such decision rules will be admissible as they are Bayes rules. Traditional justifications for using certain improper priors to induce posterior distributions have included

(i) analytic tractability

(ii) invariance arguments

(iii) the fact that such procedures yield answers which agree with answers obtained from likelihood methods.

Classically, these posterior distributions are then used to define point estimates or testing procedures as if they were defined by proper priors. That some proposed improper priors yield rather questionable solutions to statistical problems is well documented. For example, Stein (1956) provided an example where the resulting statistical procedure was uniformly inadmissible. The examples in Dawid, Stone and Zidek (1973) demonstrate that the acritical use of improper priors can lead to rather disturbing marginalization problems. However, the recent work of Sudderth (1981) has clarified the marginalization situation.

Within the context of FBDP's there is a fairly natural way of evaluating improper priors. Suppose $\gamma$ is an improper prior on $(\Theta,\mathcal{B})$ and $\gamma$ defines a posterior $Q_\gamma$ (see Section 2 for a discussion). Since $Q_\gamma$ is an inference, it can be compared to all other inferences via its risk function. For example, we can ask whether or not $Q_\gamma$ is minimax or admissible. What we are proposing is that $\gamma$ be evaluated on the basis of the risk function $R(Q_\gamma,\cdot)$. In particular, if $Q_\gamma$ is inadmissible for a variety of FBDP's, this would suggest that the use of $\gamma$ as a prior is inappropriate. On the other hand, if $Q_\gamma$ could be shown to be minimax and admissible in a number of problems, perhaps $\gamma$ could be used for other problems of a similar nature. Thus, we have a fairly well defined problem - namely, to discover conditions under which $Q_\gamma$ is admissible or inadmissible.

Although our remarks have been directed at $Q_\gamma$'s which arise from improper priors via a manipulation resulting in a posterior distribution, the same remarks apply to any other argument (e.g., the pivotal argument) which leads to a conditional distribution on $\Theta$ (that is, an inference).

Here is a brief synopsis of the remainder of this paper. In the next section, we detail our technical assumptions and describe a large class of FBLF's. This leads to the formal definition of a FBDF. Section 3 contains a discussion of a class of FBLF's called quadratic loss functions. For such loss functions, we also discuss sufficient conditions for admissibility. In Section 4, we present a few results for translation problems. In particular it is shown that for the normal distribution the "Pitman posterior" in one-dimensional translation problems is admissible and minimax. There are a number of problems of interest which do not fit directly into the framework described in Section 2. A modification of this framework is proposed in Section 5 so that marginal estimation problems and prediction problems can be handled directly.

The recent Ph.D. thesis of Gatsonis (1981) contains some results which bear a similarity to the work here although the description of his problem and that given here are quite different. The loss functions used by Gatsonis are related to those introduced in Hendrickson and Buehler (1971).

## §2: Fair Bayes Decision Problems

Suppose $\mathcal{X}$ and $\Theta$ are separable metric spaces so their Borel $\sigma$-algebras, say $\mathcal{F}$ and $\mathcal{B}$, are countably generated. Assume that

$$(2.1) \qquad \mathcal{P} = \{P(\cdot|\theta)|\theta \in \Theta\}$$

is a family of probability measures defined on the sample space $(\mathcal{X},\mathcal{F})$. Let $\mathcal{M}(\mathcal{B})$ be the set of all probability measures defined on $(\Theta,\mathcal{B})$. When equipped with the weak*-topology, $\mathcal{M}(\mathcal{B})$ is a separable metric space and $\mathcal{B}^*$ will denote the Borel $\sigma$-algebra generated by the weak*-topology (see Parthasarathy (1967), Sec. II.6). It is also known that $\mathcal{B}^*$ is the smallest $\sigma$-algebra for which all of the functions $\pi \to \int_\Theta f d\pi$ are measurable where

(i) $f$ is an arbitrary bounded continuous function

or

(ii) $f$ is the indicator of any element of a collection of sets which generates $\mathcal{B}$.

For a discussion of this, see Dubins and Freedman (1964), Parthasarathy (1967) and Blackwell, Freedman and Orkin (1974). The measurable space $(\mathcal{M}(\mathcal{B}),\mathcal{B}^*)$ will be the action space for the decision problem to be studied here. As usual, a <u>decision rule</u> $\delta$ is a measurable function defined on $(\mathcal{X},\mathcal{F})$ taking values in $(\mathcal{M}(\mathcal{B}),\mathcal{B}^*)$. For reasons discussed in Section 1, decision rules will sometimes be called <u>inferences</u>.

Remark 2.1: The usual definition of a (randomized) decision rule $\delta$ (with $\Theta$ as the action space) goes as follows: $\delta : \mathcal{B} \times X \to [0,1]$ is assumed to satisfy (i) $\delta(\cdot|x) \in M(\mathcal{B})$ for each $x \in X$ and (ii) $\delta(B|\cdot)$ is assumed to be $F$ measurable for each $B \in \mathcal{B}$. That this definition coincides with our definition follows from the remarks above concerning $\mathcal{B}^*$ (see Dubins and Freedman (1964)).

Let $\mathcal{D}$ be the class of all decision rules. For $\delta \in \mathcal{D}$, we will use the notation $\delta(\cdot|x)$ to denote the value of $\delta$ at $x \in X$ so $\delta(\cdot|x) \in M(\mathcal{B})$. Recall that a loss function L is a jointly measurable non-negative function defined on $M(\Theta) \times \Theta$. The extension of L to $M(\Theta) \times M(\Theta)$, is

$$(2.2) \qquad \bar{L}(\nu,\pi) \equiv \int_{\Theta} L(\nu,\theta)\pi(d\theta) .$$

Let $\epsilon_\theta \in M(\mathcal{B})$ denote the probability measure concentrated at $\theta \in \Theta$. Using this notation, $\bar{L}(\nu,\epsilon_\theta) = L(\nu,\theta)$. With the discussion in Section 1 as motivation, we make the following

Definition 2.1: The loss function L is called a fair Bayes loss functin (FBLF) if $\bar{L}(\nu,\pi) \geq \bar{L}(\pi,\pi)$ for all $\nu,\pi \in M(\Theta)$.

Here is an interesting class of FBLF's.

Example 2.1: Consider a jointly measurable function K on $\Theta \times \Theta$ to $R^1$ such that $K(\theta,\eta) = K(\eta,\theta)$ and

$$(2.3) \qquad \iint K(\theta,\eta)\mu(d\theta)\mu(d\eta) \geq 0$$

for all bounded signed measures $\mu$. (For example, if $f_1,\ldots,f_k$ are bounded measurable functions, then $K(\theta,\eta) = \Sigma f_i(\theta)f_i(\eta) = K(\eta,\theta)$ and (2.3) obviously holds.) Given such a $K$, define $<\cdot,\cdot>$ by

$$<\mu_1,\mu_2> = \iint K(\theta,\eta)\mu_1(d\theta)\mu_2(d\eta)$$

for bounded signed measures $\mu_1$ and $\mu_2$.

<u>Proposition 2.1</u>: Given $<\cdot,\cdot>$, define $L$ by

$$L(\nu,\theta) = <\nu-\epsilon_\theta,\nu-\epsilon_\theta>.$$

Then $L$ is a FBLF.

<u>Proof</u>: Using the bilinearity and symmetry of $<\cdot,\cdot>$ compute as follows:

$$\bar{L}(\nu,\pi) = \int<\nu-\epsilon_\theta,\nu-\epsilon_\theta>\pi(d\theta) =$$

$$\int<\nu-\pi+\pi-\epsilon_\theta,\nu-\pi+\pi-\epsilon_\theta>\pi(d\theta) =$$

$$<\nu-\pi,\nu-\pi> + \int<\pi-\epsilon_\theta,\pi-\epsilon_\theta>\pi(d\theta) =$$

$$<\nu-\pi,\nu-\pi> + \bar{L}(\pi,\pi) \geq \bar{L}(\pi,\pi) .$$

The third equality follows from the formula

$$\int<\nu,\epsilon_\theta>\pi(d\theta) = <\nu,\pi>, \quad \nu,\pi \in \mathcal{M}(\mathcal{B}) .$$

The above inequality follows from (2.3) so the proof is complete.

Loss functions of the type given in Proposition 2.1 will be called <u>quadratic loss functions</u> since $<\cdot,\cdot>$ is a positive semi-definite quadratic form.

Given any loss function L, the risk function R defined on $\mathcal{D} \times \Theta$ is

$$R(\delta,\theta) = \int_{\mathcal{X}} L(\delta(\cdot|x),\theta)P(dx|\theta) \ .$$

For $\pi \in \mathcal{M}(\mathcal{B})$, the integrated risk is

$$\bar{R}(\delta,\pi) = \int R(\delta,\theta)\pi(d\theta)$$

so $\bar{R}(\delta,\epsilon_\theta) = R(\delta,\theta)$. Of course, if $\bar{R}(\delta,\theta) \geq \bar{R}(\delta_0,\pi)$ for all $\delta$, then $\delta_0$ is a Bayes rule for $\pi$.

Given $\pi \in \mathcal{M}(\mathcal{B})$, to define the posterior distribution, first construct the joint measure on product sets by

$$\lambda(F \times B) \equiv \int_B P(F|\theta)\pi(d\theta), \quad F \in \mathcal{F}, \quad B \in \mathcal{B} \ .$$

Then, extend $\lambda$ in the obvious way to $\mathcal{F} \times \mathcal{B}$ and let

$$m_\pi(F) = \lambda(F \times \Theta)$$

be the marginal distribution on $(\mathcal{X},\mathcal{F})$. If $Q_\pi \in \mathcal{D}$ exists which satisfies

$$\lambda(F \times B) = \int_F Q_\pi(B|x)m_\pi(dx), \quad F \in \mathcal{F}, \quad B \in \mathcal{B},$$

then $Q_\pi(\cdot|x)$ is the posterior distribution on $\Theta$ given x. Sufficient conditions to guarantee the existence of $Q_\pi$ are that $\mathcal{X}$ and $\Theta$ be complete separable metric spaces (see Parthasarathy (1967), Chapter 5, Section 7). When $Q_\pi$ exists, the equation

(2.4) $$\iint h(x,\theta)P(dx|\theta)\pi(d\theta) = \iint h(x,\theta)\lambda(dx,d\theta) =$$

$$\iint h(x,\theta)Q_\pi(d\theta|x)m_\theta(dx)$$

holds for all $F \times B$ measurable $h$ which are $\lambda$-integrable.

Now, suppose that $L$ is a FBLF, $\pi \in M(B)$ and the posterior $Q_\pi$ exists.

Proposition 2.2: For $\pi \in M(B)$, the Bayes rule is $Q_\pi$.

Proof: It must be verified that

(2.5) $$\bar{R}(\delta,\pi) \geq \bar{R}(Q_\pi,\pi) \quad \text{for} \quad \delta \in \mathcal{D}.$$

Using (2.4) we have

$$\bar{R}(\delta,\pi) = \iint L(\delta(\cdot|x),\theta)P(dx|\theta)\pi(d\theta) =$$

$$\iint L(\delta(\cdot|x),\theta)Q_\pi(d\theta|x)m_\pi(dx) \geq$$

$$\iint L(Q_\pi(\cdot|x),\theta)Q_\pi(d\theta|x)m_\pi(dx) =$$

$$\iint L(Q_\pi(\cdot|x),\theta)P(dx|\theta)\pi(d\theta) =$$

$$R(Q_\pi,\pi) \ .$$

The inequality above follows from (2.2) and Definition (2.1). The proof is complete.

Definition 2.2: Any decision problem with action space $(M(B),B^*)$ for which (2.5) holds will be called a fair Bayes decision problem (FBDP).

We will now argue that FBDP's provide a structure within which methods of assigning probabilities to subsets of $\Theta$ (that is, making inferences) can be evaluated. The first point is that the rules of probability force a (proper) Bayesian to assign probabilities in a certain way - namely, via the posterior. It is appropriate to evaluate other methods of making inferences within a system which provides that Bayesians behave consistently. Any inference $\delta$ can be judged via its risk function, assuming, of course, we agree that risk functions are an appropriate measure of the behavior of $\delta$. Given this, a method of making inferences can be judged via the decision rules it produces. In particular, the use of an improper prior distribution to define an inference can be judged by the inference it produces. Of course, two properties of an inference $\delta$ which can be assessed via its risk function are admissibility and minimaxity.

To be precise, assume that $\gamma \neq 0$ is a $\sigma$-finite measure on $(\Theta,\mathcal{B})$ and $\mathcal{P} = \{P(\cdot|\theta)|\theta \in \Theta\}$ is a parametric model on $(\mathcal{X},\mathcal{F})$. Again define $\lambda$ on $\mathcal{F} \times \mathcal{B}$ by

$$\lambda(F \times B) = \iint I_F(x)I_B(\theta)P(dx|\theta)\gamma(d\theta)$$

and extend in the obvious manner so $\lambda$ is a $\sigma$-finite measure. Then, define the marginal measure on $(\mathcal{X},\mathcal{F})$ by

$$m_\gamma(F) = \lambda(F \times \Theta).$$

To proceed further, it is necessary to <u>assume</u> that $m_\gamma$ is $\sigma$-finite. This condition must be checked for each particular example. Let $\mathcal{F} \times \Theta$ denote the sub-$\sigma$-algebra of $\mathcal{F} \times \mathcal{B}$ consisting of all sets of the form $F \times \Theta$ with $F \in \mathcal{F}$. Then, the restriction of $\lambda$ to $\mathcal{F} \times \Theta$ (that is, $m_\gamma$) is $\sigma$-finite by assumption. Given a set $B \in \mathcal{B}$, the Radon-Nikodym theorem implies the existence of an $\mathcal{F} \times \Theta$ measurable function, say $Q_B$, such that

$$\int_{\mathcal{X}\times\Theta} I_F(x)Q_B(x,\theta)\lambda(d\theta,dx) = \lambda(F \times B)$$

for all $F \in \mathcal{F}$. Since $Q_B$ is $\mathcal{F} \times \Theta$ measurable, it cannot depend on the argument $\theta$ so we will write $Q_B(x)$ for $Q_B(x,\theta)$. It is easy to show $0 \leq Q_B \leq 1$ a.e. $\lambda$. Now, under suitable regularity conditions (e.g., $\mathcal{X}$ and $\Theta$ both complete separable metric spaces), we can find a decision rule $Q_\gamma$ such that for each $B \in \mathcal{B}$ $Q_\gamma(B|\cdot) = Q_B(\cdot)$ a.e. $m_\gamma$. (A proof of this parallels the proof of Theorem 8.1, Chapter 5 of Parthasarathy (1967)). This implies that

$$\iint I_F(x)I_B(\theta)Q_\gamma(d\theta|x)m_\gamma(d\theta) =$$

$$\iint I_F(x)I_B(\theta)P(dx|\theta)\gamma(d\theta), \quad F \in \mathcal{F}, \quad B \in \mathcal{B}.$$

Any such decision rule which has this property will be called a formal posterior distribution. Such posterior distributions have been proposed as decision rules. It is our suggestion that these decision rules be evaluated via their risk functions in FBDP's. In the next two sections, we look at some special cases.

## §3: Quadratic Loss Problems

In this section, we consider a decision problem with a quadratic loss function given by

$$L(\nu,\theta) = \langle \nu-\epsilon_\theta, \nu-\epsilon_\theta \rangle, \quad \nu \in \mathcal{M}(\mathcal{B})$$

where

$$\langle \mu_1,\mu_2 \rangle \equiv \iint K(\theta,\eta)\mu_1(d\theta)\mu_2(d\eta)$$

for any two bounded signed measures $\mu_1$ and $\mu_2$. Here, $K$ is a symmetric positive semi-definite kernel so $\langle \mu,\mu \rangle \geq 0$ for all bounded signed measures $\mu$ on $(\Theta,\mathcal{B})$. According to the results of the previous section, such an $L$ gives rise to a FBDP. We also make the following two simplifying assumptions:

(3.1) the kernel $K$ is bounded - say $|K(\theta,\eta)| \leq C$ for all $\theta,\eta \in \Theta$.

(3.2) for each $\delta \in \mathcal{D}$, the risk function $R(\delta,\cdot)$ is continuous.

These assumptions will make our discussion of admissibility easier. They will be verified directly for the problem of the next section.

Remark 3.1: Without assumption (3.2), the notion of almost admissibility is more appropriate to what follows.

Proposition 3.1: In order that $\delta_0$ be admissible, it is sufficient that for each non-empty open set $\mathcal{O} \subseteq \Theta$,

$$\inf_{\pi\in\Pi(\mathcal{O})} \frac{1}{\pi(\mathcal{O})} \int [R(\delta_0,\theta) - R(Q_\pi,\theta)]\pi(d\theta) = 0 \tag{3.3}$$

where $\Pi(\mathcal{O})$ is the set of $\pi \in M(\mathcal{B})$ such that $\pi(\mathcal{O}) > 0$.

Proof: This result is a minor modification of Stein's conditions (see Stein (1955)), but the proof is easy so it is included. Suppose $\delta_0$ is not admissible. Then, there exists a $\delta_1$ such that $R(\delta_1,\theta) \leq R(\delta_0,\theta)$ for all $\theta \in \Theta$, with strict inequality for some $\theta \in \Theta$. The continuity assumption on $R(\delta,\cdot)$ implies there exists an $\epsilon > 0$ such that

$$\mathcal{O} = \{\theta | R(\delta_1,\theta) < R(\delta_0,\theta) - \epsilon\}$$

is open and non-empty. Then, for each $\pi \in \Pi(\mathcal{O})$, we have

$$\frac{1}{\pi(\mathcal{O})} \int [R(\delta_0,\theta) - R(Q_\pi,\theta)]\pi(d\theta) =$$

$$\frac{1}{\pi(\mathcal{O})} \int [R(\delta_0,\theta) - R(\delta_1,\theta)]\pi(d\theta) + \frac{1}{\pi(\mathcal{O})} \int [R(\delta_1,\theta) - R(Q_\pi,\theta)]\pi(d\theta) \geq$$

$$\frac{1}{\pi(\mathcal{O})} \int_{\mathcal{O}} [R(\delta_0,\theta) - R(\delta_1,\theta)]\pi(d\theta) > \epsilon$$

which contradicts (3.3). The proof is complete.

The following result provides a useful first step in trying to verify (3.3). If $\mu$ is any bounded signed measure, $\|\mu\|^2$ will mean $\langle\mu,\mu\rangle$ so $\|\cdot\|$ is a semi-norm.

Proposition 3.2: For any quadratic loss function,

$$\int_\Theta [R(\delta,\theta) - R(Q_\pi,\theta)]\pi(d\theta) = \int_X \|\delta(\cdot|x) - Q_\pi(\cdot|x)\|^2 m_\pi(dx) \tag{3.4}$$

where $m_\pi$ is the marginal distribution on $X$.

Proof: The identity $\int\langle\mu,\epsilon_\theta\rangle\pi(d\theta) = \langle\mu,\pi\rangle$ is used below. The proof is a calculation similar to that given in the proof of Proposition 2.2. For $\pi \in M(\Theta)$, we have

$$(3.5)\quad \Psi(\pi,\delta) = \int[R(\delta,\theta) - R(Q_\pi,\theta)]\pi(d\theta) =$$

$$\iint[\|\delta(\cdot|x)-\epsilon_\theta\|^2 - \|Q_\pi(\cdot|x)-\epsilon_\theta\|^2]P_\theta(dx)\pi(d\theta) =$$

$$\iint[\langle\delta(\cdot|x),\delta(\cdot|x)\rangle-2\langle\delta(\cdot|x)-Q_\pi(\cdot|x),\epsilon_\theta\rangle-\langle Q_\pi(\cdot|x),Q_\pi(\cdot|x)\rangle]Q_\pi(d\theta|x)m_\pi(dx) =$$

$$\int[\langle\delta(\cdot|x),\delta(\cdot|x)\rangle - 2\langle\delta(\cdot|x)-Q_\pi(\cdot|x),Q_\pi(\cdot|x)\rangle - \langle Q_\pi(\cdot|x),Q_\pi(\cdot|x)\rangle]m_\pi(dx) =$$

$$\int\|\delta(\cdot|x) - Q_\pi(\cdot|x)\|^2 m_\pi(dx) .$$

This completes the proof.

To show $\delta_0$ is admissible, it suffices to show that for each open set $O$,

$$(3.6)\qquad \inf_{\pi\in\Pi(O)} \frac{\Psi(\pi,\delta_0)}{\pi(O)} = 0 .$$

There are a couple of upper bounds on $\Psi(\pi,\delta_0)$ which may be of use in the verification of (3.6).

Proposition 3.3: Given $\delta$ and $\pi$, assume that for all $x \in X$, $\delta(\cdot|x)$ and $Q_\pi(\cdot|x)$ are both absolutely continuous with respect to a fixed probability measure $\xi$ on $\Theta$. Let $\alpha(\cdot|x)$ and $q_\pi(\cdot|x)$ be the densities of $\delta(\cdot|x)$ and $Q_\pi(\cdot|x)$ with respect to $\xi$. If $|K(\theta,\eta)| \le C$ for all

$\theta, \eta \in \Theta$, then

(3.7) $$\Psi(\pi,\delta) \leq C\int_X [\int_\Theta |\alpha(\theta|x) - q_\pi(\theta|x)|\xi(d\theta)]^2 m_\pi(dx) \leq$$

$$C\int_X \int_\Theta (\alpha(\theta|x) - q_\pi(\theta|x))^2 \xi(d\theta) m_\pi(dx) .$$

Proof: Using Proposition (3.2) and the definition of $\|\cdot\|$, we have

$$\Psi(\pi,\delta) = \int_X \|\delta(\cdot|x) - Q_\pi(\cdot|x)\|^2 m_\pi(dx) =$$

$$\int[\int\int K(\theta,\eta)[\delta(d\theta|x)-Q_\pi(d\theta|x)][\delta(d\eta|x)-Q_\pi(d\eta|x)]]m_\pi(dx) =$$

$$\int[\int\int K(\theta,\eta)[\alpha(\theta|x)-q_\pi(\theta|x)][\alpha(\eta|x)-q_\pi(\eta|x)]\xi(d\theta)\xi(d\eta)]m_\pi(dx) \leq$$

$$C\int[\int\int|\alpha(\theta|x)-q_\pi(\theta|x)|\ |\alpha(\eta|x)-q_\pi(\eta|x)|\xi(d\theta)\xi(d\eta)]m_\pi(dx) =$$

$$C\int[\int|\alpha(\theta|x) - q_\pi(\theta|x)|\xi(d\theta)]^2 m_\pi(dx)$$

which is the first inequality. Applying Cauchy-Schwarz yields the second inequality so the proof is complete.

Another upper bound for $\Psi(\pi,\delta)$ can be given in terms of the Hellinger distance (see Kakutani (1948)). Given two probability measures $\pi_1$ and $\pi_2$, the squared Hellinger distance is given by

(3.8) $$H^2(\pi_1,\pi_2) = \int\left[\left(\frac{d\pi_1}{d\beta}\right)^{\frac{1}{2}} - \left(\frac{d\pi_2}{d\beta}\right)^{\frac{1}{2}}\right]^2 d\beta$$

where $\beta$ is any measure which dominates both $\pi_1$ and $\pi_2$. Of course, the value of $H(\pi_1,\pi_2)$ does not depend on the choice of $\beta$. For $\pi_1,\pi_2 \in M(\Theta)$, let $\alpha_i$ be the density of $\pi_i$ with respect to $\beta$. Then

(3.9) $$\|\pi_1 - \pi_2\|^2 =$$

$$\iint K(\theta,\eta)[\alpha_1(\theta)-\alpha_2(\theta)][\alpha_1(\eta)-\alpha_2(\eta)]\beta(d\theta)\beta(d\eta) \leq$$

$$C[\int|\alpha_1(\theta)-\alpha_2(\theta)|\beta(d\theta)]^2 =$$

$$C[\int|\alpha_1^{\frac{1}{2}}(\theta)-\alpha_2^{\frac{1}{2}}(\theta)|\,|\alpha_1^{\frac{1}{2}}(\theta)+\alpha_2^{\frac{1}{2}}(\theta)|\beta(d\theta)]^2 \leq$$

$$CH^2(\pi_1,\pi_2)\int(\alpha_1^{\frac{1}{2}}(\theta)+\alpha_2^{\frac{1}{2}}(\theta))^2\beta(d\theta) =$$

$$2CH^2(\pi_1,\pi_2)[1 + \int\alpha_1^{\frac{1}{2}}(\theta)\alpha_2^{\frac{1}{2}}(\theta)\beta(d\theta)] \leq$$

$$4CH^2(\pi_1,\pi_2) .$$

This leads to

Proposition 3.4: Given $\delta$ and $\pi$,

(3.10) $$\Psi(\pi,\delta) \leq 4C\int H^2(\delta(\cdot|x),Q_\pi(\cdot|x))m_\pi(dx) .$$

Proof: Apply (3.9) with $\pi_1 = \delta(\cdot|x)$, $\pi_2 = Q_\pi(\cdot|x)$ and use Proposition 3.2.

An upper bound similar to (3.10) was given by Stein (1965) in his study of admissibility in classical decision theory problems. This suggests that admissibility in the present context will be closely connected to admissibility in more classical problems - at least in problems where Stein's upper bound is valid.

## §4: A Simple Example

Our discussion in this section will be centered around the estimation of a univariate normal mean. Unfortunately, space limitations prohibit the inclusion of some of the detailed calculations. Suppose $X$ is $N(\theta,1)$ so $\mathcal{X} = \Theta = R^1$ ( a sufficiency argument shows we need only consider the case of one observation). Consider the particular inference $\delta_0$ which specifies that $\theta$ is $N(x,1)$ when $X = x$ - that is, $\delta(\cdot|x)$ is the $N(x,1)$ distribution on $R^1$. To define a FBLF, consider the kernel

$$K_t(\theta,\eta) = \exp[it(\theta-\eta)]$$

for $t \in R^1$ and $\theta,\eta \in R^1$. Note that $K_t$ is complex valued. It is easy to verify that the bilinear form $\langle\cdot,\cdot\rangle$ defined on pairs of bounded signed measures by

$$\langle\mu_1,\mu_2\rangle_t = \iint K_t(\theta,\eta)\mu_1(d\theta)\mu_2(d\eta)$$

is positive semi-definite. The calculation given in Example 2.1 shows that the loss function $L_t$ defined by

$$L_t(\nu,\theta) = \langle\nu-\epsilon_\theta,\nu-\epsilon_\theta\rangle_t$$

is a FBLF for each $t$. Further, the continuity of the risk function $R_t(\delta,\cdot)$ for each decision rule is easily verified using the analytic properties of the $N(\theta,1)$ distribution.

Given $\sigma > 0$, let $\pi_\sigma$ denote the $N(0,\sigma^2)$ distribution and let $Q_\sigma$ denote the posterior distribution of $\theta$ when $X$ is $N(\theta,1)$. Thus $Q_\sigma(\cdot|x)$ is a $N(cx,c)$ distribution where $c = \sigma^2/(1+\sigma^2)$. Also, the marginal distribution of $X$ is $N(0,1+\sigma^2)$ which is denoted by $m_\sigma$. Let $\|\cdot\|_t$ denote the semi-norm defined by $K_t$.

Lemma 4.1: For each non-empty open set $\mathcal{O} \subseteq R^1$,

$$(4.1) \qquad \lim_{\sigma\to\infty} \sup_t (\pi_\sigma(\mathcal{O}))^{-1} \int \|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|_t^2 m_\sigma(dx) = 0 .$$

Proof: Since each $\mathcal{O}$ contains an interval, it is sufficient to establish (4.1) where $\mathcal{O} = (a,b)$, $-\infty < a < b < \infty$. In this case, note that $\sigma\pi_\sigma(\mathcal{O}) \to (b-a)$ as $\sigma \to \infty$ so it suffices to verify (4.1) with $(\pi_\sigma(\mathcal{O}))^{-1}$ replaced by $\sigma$. From the special form of $K_t$, it is clear that $\|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|_t^2$ is the modulus squared of the difference between the characteristic functions of $\delta_0(\cdot|x)$ and $Q_\sigma(\cdot|x)$ evaluated at $t$. Since the two distributions are normal, a calculation yields

$$(4.2) \qquad \Delta_\sigma(x,t) = \|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|_t^2 = |\exp(itx-\tfrac{1}{2}t^2) - \exp(ictx-\tfrac{1}{2}c^2t^2)|^2$$

where $c = \sigma^2/(1+\sigma^2)$. Integrating $\Delta_\sigma(x,t)$ with respect to $m_\sigma(dx)$ yields

$$(4.3) \qquad \sup_t \int \Delta_\sigma(x,t) m_\sigma(dx) =$$

$$\sup_t \left\{\exp[-t^2] + \exp[-c^2t^2] - 2\exp\left[-t^2\left(1 + \frac{c(c-1)}{2}\right)\right]\right\} \le$$

$$\sup_t \left\{|\exp[-t^2] - \exp[-c^2t^2]| + 2\left|\exp\left[-t^2\left(1 + \frac{c(c-1)}{2}\right)\right] - \exp[-c^2t^2]\right|\right\} \le$$

$$3\sup_t |\exp[-t^2] - \exp[-c^2t^2]| + 2\sup_t \left|\exp\left[-t^2\left(1 + \frac{c(c-1)}{2}\right)\right] - \exp[-t^2]\right| .$$

Now, it is a routine but tedious calculation to compute the two suprema in the final term of (4.3). Using this result, one then can show that

$$\lim_{\sigma\to\infty} \sigma \sup_t \int \Delta_\sigma(x,t) m_\sigma(dx) = 0$$

which completes the proof.

Proposition 4.1: Let $k$ be a complex valued function defined on $R^1$ such that $k(0) = 1$ and the kernel $K$ given by $K(\theta,\eta) = k(\theta-\eta)$ satisfies $K(\theta,\eta) = \overline{K(\theta,\eta)}$ where the bar denotes complex conjugate. Assume $K$ is positive semi-definite and let $L$ be the FBLF defined by $K$. Then $\delta_0$ is admissible in the FBDP defined by $L$.

Proof: The assumption that $k(0) = 1$ together with the assumption that $K$ is positive semi-definite allows us to invoke Bochner's Theorem to conclude that

$$K(\theta,\eta) = \int_{-\infty}^{\infty} K_t(\theta,\eta) H(dt)$$

where $H$ is a distribution function on $R^1$. To show $\delta_0$ is admissible, we will verify (3.3) using (3.4) and Lemma 4.1. As in the proof of Lemma 4.1, it suffices to take $\mathcal{O} = (a,b)$ where $-\infty < a < b < \infty$ and it suffices to show that

$$\lim_{\sigma\to\infty} \sigma \int \|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|^2 m_\sigma(dx) = 0, \tag{4.4}$$

where $\|\cdot\|$ is the semi-norm defined by $K$. But, for any bounded signed measure $\mu$,

$$\|\mu\|^2 = \langle\mu,\mu\rangle = \int\langle\mu,\mu\rangle_t H(dt) = \int\|\mu\|_t^2 H(dt) .$$

Substituting this into (4.4) and using Lemma 4.1 yields

$$\lim_{\sigma\to\infty} \sigma \iint \|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|_t^2 H(dt) m_\sigma(dx) \le$$

$$\lim_{\sigma\to\infty} \sigma \sup_t \int \|\delta_0(\cdot|x) - Q_\sigma(\cdot|x)\|_t^2 m_\sigma(dx) = 0.$$

This completes the proof.

We close this section with a few remarks.

Remark 4.1: For the problem above with a general bounded kernel $K$, it should be possible to show that $\delta_0$ is admissible via (3.3) and (3.4) by using the upper bounds given in Proposition 3.3 or Proposition 3.4. I have been unable to carry out the calculations thus far.

Remark 4.2: For loss functions of the type assumed in Proposition 4.1, the resulting decision problem is invariant (in the traditional sense) under translations. Of course, the group $G$ is $R^1$ and $G$ acts in the obvious way on $\mathcal{X}$ and $\Theta$. The action of $G$ on decision rules is defined by

$$(\hat{g}\delta)(B|x) = \delta(g^{-1}B|g^{-1}x), \quad g \in G$$

where $B$ is a Borel set of $\Theta$ and $x \in \mathcal{X}$. Then $\delta$ is called invariant if $\hat{g}\delta = \delta$. Since the group $G$ acts transitively on $\Theta$, any invariant decision rule will have constant risk. The best of all the invariant rules is $\delta_0$ - this follows from its admissibility, but an appeal to a general result of Stein is more appropriate (see Zidek (1969)) - namely, under conditions which hold for this problem, the best invariant rule is obtained by using the right Haar measure as a prior (improper) distribution and calculating the posterior distribution. Of course, Lebesgue measure is the right Haar measure on $G$ and $\delta_0$ is the formal posterior distribution for this improper prior. Since $\delta_0$ is a best invariant decision rule, that it is minimax follows from a result due to Kiefer (1957).

Remark 4.3: When $X$ is $N_p(\theta, I_p)$ and $p \geq 3$, it is natural to ask if the obvious decision rule, $\delta_0$, which specifies that $\theta$ given $X = x$ is $N(x, I_p)$ is admissible. One suspects not because of the existence of Stein-type estimators when $p \geq 3$. Here is what I know so far. For the kernel $K$ given by

$$K(\theta,\eta) = \exp[-\tfrac{1}{2}\|\theta-\eta\|^2]$$

which defines a FBLF and a FBDP, $\delta_0$ is minimax (Kiefer's Theorem (1957)) but $\delta_0$ is not admissible. Consider a decision rule of the form

$$\delta_1(\cdot|x) \quad \text{is} \quad N(u(x), I_p)$$

where

$$u(x) = \left(1 - \frac{a}{b+\|x\|^2}\right)x \ .$$

Then there exists an $a$ (small and positive) and a $b$ (large and positive) so that $\delta_1$ dominates $\delta_0$. Methods similar to those in Brown (1966) are used to establish this claim. The proof is tedious and offers virtually no guidance in the important problem of finding a prior (proper or improper) so that the corresponding posterior dominates $\delta_0$. A result similar to the above has recently been established by Gatsonis (1981).

## §5: Some Extensions

In this section we extend our formulation of the decision problem to include the marginal estimation of parameters problem and the prediction problem. To discuss the first problem, again assume $(X,F)$ is the sample space, $(\Theta,\mathcal{B})$ is the parameter space and a model $\{P(\cdot|\theta)|\theta \in \Theta\}$ is given. Rather than requiring an inference from $X$ to $M(\mathcal{B})$, we now consider a $\sigma$-algebra $\mathcal{B}_0 \subseteq \mathcal{B}$, and define an inference to be measurable map from $X$ to $M(\mathcal{B}_0)$ where $M(\mathcal{B}_0)$ is the set of all probability measures on $(\Theta,\mathcal{B}_0)$. An example will illustrate how this situation arises.

Example 5.1: Suppose $X_1,\ldots,X_n$ is a random sample from a $N(\mu,\sigma^2)$ population with $\mu$ and $\sigma^2$ both unknown. Thus, $X = R^n$ and $\Theta = \{\theta|\theta = (\mu,\sigma^2),\ \mu \in R^1,\ \sigma^2 > 0\}$, so $\mathcal{B}$ is the set of Borel sets of $\Theta$. If we are only interested in inferences about $\mu$, then we take $\mathcal{B}_0 \subseteq \mathcal{B}$ to be

$$\mathcal{B}_0 = \{C\times(0,\infty)|C \text{ is a Borel set of } R^1\} .$$

Given any probability measure $\pi \in M(\mathcal{B})$, its projection to $M(\mathcal{B}_0)$ is given by simply restricting $\pi$ to $\mathcal{B}_0$. For this example, the projection corresponds to "integrating out $\sigma$" or "marginalizing". That is, we can define $\tilde{\pi}$ on $R^1$ by $\tilde{\pi}(C) = \pi(C\times(0,\infty))$ so $\tilde{\pi}$ is just the marginal of $\pi$ and provides an inference for $\mu \in R^1$.

To continue with the general discussion, for $\pi \in M(\mathcal{B})$, let $\hat{\pi}$ denote the restriction of $\pi$ to $\mathcal{B}_0$ so $\hat{\pi} \in M(\mathcal{B}_0)$. In what follows, we assume that all elements of $M(\mathcal{B}_0)$ are obtained by the restriction of elements of $M(\mathcal{B})$. A sufficient condition for this is that both $(\mathcal{X},\mathcal{B})$ be Polish and $\mathcal{B}_0$ be countably generated (see Ascherl and Lehn (1977) for a discussion). Given a loss function L defined on $M(\mathcal{B}_0) \times \Theta$, define $\bar{L}$ on $M(\mathcal{B}_0) \times M(\mathcal{B})$ by

$$\bar{L}(\nu,\pi) = \int L(\nu,\theta)\pi(d\theta) \ .$$

Then L is a <u>FBLF</u> if $\bar{L}(\nu,\pi) \geq \bar{L}(\hat{\pi},\pi)$ for all $\nu \in M(\mathcal{B}_0)$ and $\pi \in M(\mathcal{B})$. Any such FBLF will give rise to a risk function R and its extension $\bar{R}$. When L is a FBLF, it is easy to show that

$$\bar{R}(\delta,\pi) \geq \bar{R}(\hat{Q}_\pi,\pi) \ .$$

The proof of this is the same as given in Proposition 2.2. In other words, given a FBLF and $\pi$, the Bayes solution to the decision problem is obtained by projecting the posterior distribution. Examples of L's which are FBLF's are easily constructed as in Example 2.1 by taking the kernel K to be $\mathcal{B}_0 \times \mathcal{B}_0$ measurable. The verification of this is essentially that given for Proposition 2.1. Given such a quadratic loss function, all of the results in Section 3 are valid with virtually no modifications.

We now briefly indicate how the prediction problem can be formulated within the present context. Suppose $(\mathcal{X},\mathcal{F})$ is a sample space, $(\mathcal{Z},\mathcal{G})$ is is a space of values for future observables and $(\P,\mathcal{C})$ is a parameter space. The observation $X \in \mathcal{X}$ is assumed to have a distribution belonging to a family $\mathcal{P} = \{P(\cdot|z,\eta)\ z \in \mathcal{Z},\ \eta \in \P\}$. The future observable Z is assumed to have a distribution in the family $\mathcal{S} = \{S(\cdot|\eta)|\eta \in \P\}$. The decision problem is to make an inference about $Z$ after observing $X$. Let $(\Theta,\mathcal{B}) = (Z \times \P, \mathcal{G} \times \mathcal{C})$ be the "parameter space" for this problem. Rather than consider the set of all "prior" distributions on $(\Theta,\mathcal{B})$, our assumptions require us to look only at distributions on $(\Theta,\mathcal{B})$, say $\pi$, which have the form

$$(5.1) \qquad \pi(dz,d\eta) = S(dz|\eta)\xi(d\eta)$$

where $\xi$ is an arbitrary prior on $(\P,\mathcal{C})$ - that is, $\xi \in M(\mathcal{C})$. In other words, the model assumptions specify the conditional distribution of $z$ given $\eta$ which is exactly what (5.1) means. Let $M_0(\mathcal{G} \times \mathcal{C})$ be distributions of the form (5.1). Since the problem is to make an inference about $z$, a decision rule should be a function from $X$ to $M(\mathcal{G})$. However, it is a bit more consistent to identify $M(\mathcal{G})$ with projections of measures in $M(\mathcal{B})$. To be more precise, let $\mathcal{B}_0$ be the sub-$\sigma$-algebra of $\mathcal{B}$ defined by

$$\mathcal{B}_0 = \{G \times \P | G \in \mathcal{G}\} .$$

Thus, the restriction of $\pi \in M(\mathcal{B})$ to $\mathcal{B}_0$ defines a distribution in $M(\mathcal{G})$ and conversely. Finally, an <u>inference</u>, $\delta$, is a measurable function defined on $X$ taking values in $M(\mathcal{B}_0)$.

With the above formulation of the prediction problem, we have a problem very similar to the marginalization problem discussed earlier in the section. The only difference is that the prior distributions are of a restricted form because of our prior assumptions concerning the model. It is now a routine matter to extend the earlier notions. A loss function $L(\delta,\theta)$ is a FBLF if for $\pi \in M_0(\mathcal{G} \times \mathcal{C})$,

$$\int L(\delta,\theta)\pi(d\theta) \geq \int L(\hat{\pi},\theta)\pi(d\theta)$$

where $\hat{\pi}$ is the restriction of $\pi$ to $\mathcal{B}_0$ ($\hat{\pi}$ is the marginal of $\pi$ on $(\mathcal{Z},\mathcal{G})$). Given such a FBLF, the risk function R will have the property that

$$(5.2) \qquad \int R(\delta,\theta)\pi(d\theta) \geq \int R(\hat{Q}_\pi,\theta)\pi(d\theta)$$

for $\pi \in \mathcal{M}_0(\mathcal{G}\times\mathcal{C})$. Here, $\hat{Q}_\pi$ is the projection of the posterior $Q_\pi$ onto $\mathcal{B}_0$. In other words, given a $\pi$ of the form (5.1) and $x \in \mathcal{X}$, we have a posterior distribution, say $Q_\pi(\cdot|x)$ defined on $(\mathcal{Z}\times\P,\mathcal{G}\times\mathcal{C})$. Then, the projection of $Q_\pi(\cdot|x)$ is defined by

$$\hat{Q}_\pi(G\times\P|x) \equiv Q_\pi(G\times\P|x)$$

for each set $G \in \mathcal{G}$ – that is, $\hat{Q}_\pi(\cdot|x)$ is the marginal distribution of $Q_\pi(\cdot|x)$ on the space $(\mathcal{Z},\mathcal{G})$. Of course, $\hat{Q}_\pi(\cdot|x)$ is ordinarily called the predictive distribution. The above argument shows that if L is a FBLF, then given $\pi \in \mathcal{M}_0(\mathcal{G}\times\mathcal{C})$, the Bayes solution to the decision problem (the prediction problem) is just the predictive distribution. As with the marginalization problem described earlier, examples of FBLF's are provided by $\mathcal{B}_0\times\mathcal{B}_0$ measurable kernels which are positive semi-definite (as argued in Example 2.1).

## References

[1] Ascherl, A. and Lehn, J. (1977). Two principles for extending probability measures. Manuscripta Math., 21, p. 43-50.

[2] Bernardo, J. M. (1979). Expected information as expected utility. Ann. Statist., 7, p. 686-690.

[3] Blackwell, D. (1956). On a class of probability spaces. Proc. 3rd Berkeley Symp. Math. Statist. Probab., Vol II, p. 1-6.

[4] Blackwell, D., Freedman, D. and Orkin, M. (1974). The optimal reward operator in dynamic programming. Ann. Probab., 2, p. 926-941.

[5] Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Rev., 78, p. 1-3.

[6] Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Statist., 37, p. 1087-1131.

[7] Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. J. R. Statist. Soc. B, 35, p. 189-233.

[8] Dubins, L. and Freedman, D. (1964). Measurable sets of measures. Pac. J. Math., 14, p. 1211-1222.

[9] Gatsonis, C. A. (1981). Estimation of the posterior density of a location parameter. Ph.D. Thesis, Cornell University.

[10] Good, I. J. (1952). Rational decisions. J. R. Statist. Soc. B, 14, p. 357-364.

[11] Hendrickson, A. D. and Buehler, R. J. (1971). Proper scores for probability forecasters. Ann. Math. Statist., 42, p. 1916-1921.

[12] Kakutani, S. (1948). On equivalence of infinite product measures. Ann. Math., 49, p. 214.

[13] Kiefer, J. C. (1957). Invariance, minimax sequential estimation and continuous time processes. Ann. Math. Statist., 28, p. 573-601.

[14] Parthasarathy, K. R. (1967). Probability Measures on Metric Spaces, Academic Press, New York.

[15] Savage, L. J. (1971). The elicitation of personal probabilities and expectations. J. Amer. Statist. Assn., **66**, p. 783-801.

[16] Stein, C. (1955). A necessary and sufficient condition for admissibility. Ann. Math. Statist., **26**, p. 518-522.

[17] Stein, C. (1956). Some problems in multivariate analysis, Part I. Stanford University Technical Report No. 6.

[18] Stein, C. (1965). Approximation of improper prior measures by prior probability measures. Bernoulli, Bayes, Laplace Anniversary Volume, ed. by J. Neyman and L. LeCam, p. 217-240. Springer-Verlag, Berlin.

[19] Sudderth, W. (1981). Finitely additive priors, coherence and the marginalization paradox. J. Roy. Statist. Soc. B, **42**, p. 339-341.

[20] Zidek, J. V. (1969). A representation of Bayes invariant procedures in terms of Haar measure. Ann. Inst. Statist. Math., **22**, p. 291-308.